COMPARATIVE EFFICACY AND ACCEPTABILITY OF ANTIDEPRESSANT DRUGS IN THE ACUTE TREATMENT OF MAJOR DEPRESSIVE DISORDER: A NETWORK META-ANALYSIS

Andrea Cipriani,[1] Toshi A Furukawa,[2§] Georgia Salanti,[3§] Anna Chaimani,[4] Lauren Z Atkinson,[1] Yusuke Ogawa,[2] Stefan Leucht,[5] Henricus G Ruhe,[1,6] Erick H Turner,[7] Julian PT Higgins,[8] Matthias Egger,[3] Nozomi Takeshima,[2] Yu Hayasaka,[2] Hissei Imai,[2] Kiyomi Shinohara,[2] Aran Tajika,[2] John PA Ioannidis,[9] John R Geddes[1]

[1] Department of Psychiatry, University of Oxford, Oxford, United Kingdom; Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, United Kingdom.
[2] Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan.
[3] Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland.
[4] Paris Descartes University, Paris, France; INSERM, UMR1153 Epidemiology and Statistics, Sorbonne Paris Cité Research Center (CRESS), METHODS Team, Paris, France; Cochrane France, Paris, France.
[5] Department of Psychiatry and Psychotherapy, TU-Munich, Munich, Germany.
[6] Department of Psychiatry, Radboudumc Nijmegen, The Netherlands.
[7] Behavioral Health and Neurosciences Division, VA Portland Health Care System, Portland, Oregon, United States of America & Departments of Psychiatry and Pharmacology, Oregon Health & Science University, Portland, Oregon, United States of America.
[8] School of Social and Community Medicine, University of Bristol, Bristol, UK.
[9] Department of Medicine, Stanford University School of Medicine, Stanford, USA; Department of Health Research and Policy, Stanford University School of Medicine, Stanford, USA; Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, USA; Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, USA; Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, USA.
[§] Joint first authors.

Correspondence to:

Andrea Cipriani, MD PhD
Department of Psychiatry
Warneford Hospital
Oxford, OX3 7JX
Email: andrea.cipriani@psych.ox.ac.uk

Abstract

Background. Major depressive disorder is one of the most common, burdensome and costly psychiatric disorders worldwide in adults. Both pharmacological and non-pharmacological treatments are available, however, because of lack of resources, antidepressants are used more frequently. Prescription of these agents should be informed by the best available evidence. Consequently, we aimed to update and expand our previous work to compare and rank antidepressants for major depressive disorder in adults.

Methods We searched Cochrane CENTRAL, CINAHL, EMBASE, LiLACS, MEDLINE, PSYCINFO, regulatory agencies' websites, and international registers for published and unpublished, double-blind randomised controlled trials up to January 8th 2016, for the acute treatment of major depressive disorder diagnosed according to standard operationalised criteria. We included placebo-controlled and head-to-head trials of 21 antidepressants in adults. We assessed the certainty of evidence using GRADE. Primary outcomes were efficacy (response rate) and acceptability (discontinuations due to any cause). Secondary outcomes included symptom severity, remission rate and discontinuations due to adverse events. We estimated summary odds ratios (OR) and standardised mean differences (with 95% credibility intervals - 95% CrIs) using pairwise and network meta-analysis with random effects. This study is registered with PROSPERO (CRD42012002291).

Findings. We included 522 trials with 116,477 participants. The certainty of evidence was moderate to very low. In terms of efficacy, all antidepressants were more effective than placebo, with OR ranging between 2·13 (95% CrI 1·89 to 2·41) for amitriptyline and 1·38 (95% CrI 1·16 to 1·63) for reboxetine. For acceptability, agomelatine and fluoxetine were associated with fewer dropouts than placebo (OR 0·84, 95% CrI 0·72 to 0·97 and 0·88, 95% CrI 0·80 to 0·96, respectively), while clomipramine was worse than placebo (OR 1.31, 95% CrI 1·01 to 1·68). When all trials were considered, differences in OR between antidepressants ranged from 1·15 (95% CrI 1·04 to 1·27) to 1·55 (95% CrI 1·27 to 1·91) for efficacy and from 0.64 (95% CrI 0·48 to 0·86) to 0.85 (95% CrI 0·75 to 0·96) for acceptability, with wide confidence intervals on most of the comparative analyses.

In head-to-head studies, agomelatine, amitriptyline, escitalopram, mirtazapine, paroxetine, sertraline, venlafaxine and vortioxetine were more effective than other antidepressants (OR range: 1.12 [95% CrI 1·00 to 1·32] to 1.96 [95% CrI 1·09 to 3·57]), while fluoxetine, reboxetine and trazodone were the least efficacious drugs (OR range: 0.51 [95% CrI 0·72 to 0·97] to 0.89 [95% CrI 0·72 to 0·97]). For acceptability, agomelatine, citalopram, escitalopram, fluoxetine, sertraline and vortioxetine were the best drugs (OR range: 0.42 [95% CrI 0·72 to 0·97] to 0.81 [95% CrI 0·72 to 0·97]), while amitriptyline, clomipramine, duloxetine, fluvoxamine, reboxetine, trazodone and venlafaxine had the highest dropout rates (OR range: 1.23 [95% CrI 1·00 to 1·32] to 2.37 [95% CrI 1·00 to 1·32]).

Interpretation. All antidepressants were more efficacious than placebo in adults with major depressive disorder. Smaller differences between active drugs were found when placebo-controlled trials were included in the analysis, while there was more variability in efficacy and rate of drop out in head-to-head trials. These results should serve evidence-based practice and inform patients, physicians, guideline developers and policy-makers on the relative merits of the different antidepressants.

Funding. NIHR Oxford Health Biomedical Research Centre

Research in context

Evidence before this study

Antidepressants are used worldwide for the treatment of major depressive disorder; however, in the scientific literature there is a long-lasting debate about their effectiveness and the potential differences in profile of the individual drugs. The aim of this systematic review and network meta-analysis is to better inform clinical practice and mental health policies by comparing all licensed second-generation as well as four reference first-generation antidepressants against each other and with placebo in terms of efficacy and acceptability for the acute treatment of adults with unipolar major depressive disorder.

Added value of this study

This study is the update and extension of our previous network meta-analysis that addressed 12 antidepressants with data on head-to-head comparisons only. In the present analysis, we managed to include 522 double-blind randomised controlled trials (116,477 patients) conducted between 1979 and 2016, and comparing 21 antidepressants (agomelatine, bupropion, citalopram, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, paroxetine, reboxetine, sertraline, venlafaxine, vilazodone, and vortioxetine) or placebo. All antidepressants were more efficacious than placebo, and some antidepressants were possibly better in terms of efficacy and/or acceptability than other active drugs. Estimated differences between antidepressants were smaller in placebo-controlled trials compared to head-to-head studies.

Implications of all the available evidence

Findings from this network meta-analysis represent the best currently available evidence-base to guide the choice about pharmacological treatment for acute major depressive disorder in adults. This evidence should

inform clinical guidelines and assist the shared decision-making process between patients, carers and clinicians in routine practice.

Introduction

Psychiatric disorders account for 22.8% of the global burden of diseases.[1] The leading cause of disability is depression, which has dramatically increased since 1990, largely driven by population growth and aging.[2] With an estimated 350 million people affected across the world, the economic burden of depressive disorders in the US alone has been estimated to be more than $210 billion, with approximately 45% attributable to direct costs, 5% to suicide-related costs, and 50% to workplace costs.[3] This poses a significant challenge for health systems in both developed and developing regions, with the need to treat patients, optimise resources and improve overall healthcare in mental health. Grouped into classes of drugs with slightly different mechanisms of action, antidepressants are widely used treatments for major depressive disorder, which are available worldwide. However, there is a long-lasting debate and concern about their efficacy and effectiveness, as short-term benefits are, on average, modest and long-term balance of benefits and harms is often understudied.[4] Innovation in psychopharmacology is therefore of crucial importance, but the identification of new molecular targets is difficult, primarily due to the lack of knowledge about how antidepressants work.[5] In routine practice, clinicians have a wide choice of individual drugs and they need good evidence to make the best choice for each individual patient. Network meta-analyses of existing datasets makes it possible to estimate comparative efficacy, summarise and interpret the wider picture of the evidence base, and to understand the relative merits of the multiple interventions.[6]

The aim of this systematic review is to inform clinical practice by comparing antidepressants in the acute treatment of adults with unipolar major depressive disorder. The project extends our previous work that had addressed 12 antidepressants with data on head-to-head comparisons.[7] The present analysis is substantially more comprehensive because it includes 21 active treatments and placebo, in addition to three new clinical outcome measures.

METHODS

Search strategy and selection criteria

The full protocol of this network meta-analysis was registered with PROSPERO (CRD42012002291) and published.[8] We searched the Cochrane Central Register of Controlled Trials, CINAHL, EMBASE, LiLACS, MEDLINE, MEDLINE In-Process and PSYCINFO from the date of database inception to January 8th 2016. The search criterion was limited to double-blind randomised controlled trials (RCTs) comparing antidepressants with placebo or another active antidepressant as oral monotherapy in the acute treatment of adults (≥18 years old) with a primary diagnosis of major depressive disorder according to standard operationalised diagnostic criteria, including Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, DSM-5 and ICD-10. We considered only double-blind trials because we included placebo in the network and because this study design increases methodological rigour by minimising performance and ascertainment biases.[9] We put no restrictions on language. The electronic database search was supplemented with manual searches for published, unpublished and ongoing RCTs in international trial registers, websites of drug-approval agencies and key scientific journals in the field.[8] We contacted all the pharmaceutical companies marketing antidepressants and asked for supplemental unpublished information on both premarketing and postmarketing studies, with a specific focus on second-generation antidepressants. Study authors and drug manufacturers were contacted to supplement incomplete reports of the original papers or provide data for unpublished studies.

We included all second-generation antidepressants approved by the regulatory agencies in USA or in Europe: agomelatine, bupropion, citalopram, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, paroxetine, reboxetine, sertraline, venlafaxine, vilazodone, and vortioxetine. To inform clinical practice globally, we selected the two tricyclics included in the WHO Model List of Essential Medicines, namely amitriptyline and clomipramine (http://www.who.int/medicines/publications/essentialmedicines/20th_EML2017.pdf?ua=1). We also included trazodone and nefazodone, because of their distinct effect and tolerability profiles. Trials including

20% or more of participants with bipolar disorder, psychotic or treatment-resistant depression or patients with a serious concomitant medical illness were excluded. Trials which allowed rescue medications were included so long as they were equally provided among the randomised arms. We included data only on drugs within the therapeutic range (Appendix 4.4).

Data extraction and quality assessment

Six pairs of investigators (ACi, TAF, LZA, SL, HGR, YO, NT, YH, EHT, HI, KS, AT) independently selected studies, reviewed the main reports and supplementary materials, extracted the relevant information from the included trials, and assessed the risk of bias with the Cochrane tool (http://handbook-5-1.cochrane.org/). Any discrepancies were resolved by consensus and arbitration by a panel of investigators within the review team (ACi, TAF, LZA, EHT, JRG).

Outcomes

Our primary outcomes were *efficacy* (response rate measured by the total number of patients who had a reduction of at least 50% of the total score on a standardised observer-rating scale for depression) and *acceptability* (treatment discontinuation measured by the proportion of patients who withdrew for any reason).[8] All-cause discontinuation was used as a measure for the acceptability of treatments, because it encompasses efficacy and tolerability.[7] Secondary outcomes included endpoint depression score, remission rate and the proportion of patients who dropped out early due to adverse events. When depressive symptoms had been measured with more than one standardised rating scale, we used a predefined hierarchy, based on psychometric properties and consistency of use across included trials.[8] In the absence of information or supplemental data from the authors, response rate was calculated according to a validated imputation method.[10] We recorded the outcomes as close to 8 weeks as possible for all analyses.[7] If information at 8 weeks was not available, we used data ranging between 4 and 16 weeks (we gave preference to the time point closest to 8 weeks; if equidistant, we took the longer outcome). We checked trial protocols

where available and compared published and unpublished data. We extracted data following a pre-defined hierarchy that gave priority to unpublished information in case of disagreement.[8]

Statistical analysis

Full details of the applied statistical approaches are provided in the protocol.[8] We estimated summary odds ratios (OR) for dichotomous outcomes and standardised mean differences (SMD, Cohen's d) for continuous outcomes using pairwise and network meta-analysis. In network meta-analysis we used arm-level data; the binomial likelihood was used for dichotomous outcomes and the normal likelihood for continuous outcomes. The study effect sizes were then synthesised using a random effects network meta-analysis model. We accounted for the correlations induced by multi-arm studies by employing multivariate distributions. The variance in the random-effects distribution (heterogeneity variance) was considered to measure the extent of across-study and within-comparison variability on treatment effects. In network meta-analysis we assumed that the amount of heterogeneity was the same for all treatment comparisons. To assess the amount of heterogeneity, we compared the posterior distribution of the estimated heterogeneity variance with its predictive distribution.[11] To rank the treatments for each outcome, we used the surface under the cumulative ranking curve (SUCRA) and the mean ranks.[12] The transitivity assumption underlying network meta-analysis was evaluated by comparing the distribution of clinical and methodological variables that could act as effect modifiers across treatment comparisons.[8] We performed a statistical evaluation of consistency (i.e. the agreement between direct and indirect evidence) using the design-by treatment test [13] and by separating direct from indirect evidence.[14]

We explored whether treatment effects for the two primary outcomes were robust in subgroup analyses and network meta-regression using the following characteristics: (1) study year; (2) sponsorship; (3) depressive severity at baseline; (4) dosing schedule; (5) study precision (i.e. small study effect); and (6) novelty effect.[15] See Appendix 4 for the definition of covariates. The sensitivity of our conclusions was evaluated by analysing the dataset with the following restrictions: (1) studies with reported response rate; (2) studies using accepted doses in all arms; (3) studies with unpublished data; (4) multi-centre studies; and (5) head-to-head studies.

We used comparison-adjusted funnel plots to investigate whether results in imprecise trials differ from those in more precise trials.[16]

We fitted all models in OpenBUGS [17] using the binomial likelihood for dichotomous outcomes, uninformative prior distributions for the treatment effects and a minimally informative prior distribution for the common heterogeneity standard deviation. We assumed uninformative priors, that is N(0,1000), for all meta-regression coefficients. Convergence of models was ensured by visual inspection of three chains and after considering the Brooks-Gelman-Rubin diagnostic. The codes of analyses, statistical details of the meta-analysis and meta-regression models are presented in Appendix 7.6.1. Statistical evaluation of inconsistency and production of network graphs and result figures were carried out using the network and network graphs packages in Stata.[18] Network meta-analyses of the primary outcomes were duplicated using the *netmeta* package in R.[19] We assessed the certainty of evidence contributing to network estimates of the main outcomes with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.[20]

Appendix 10 lists the changes to the original protocol. The study was conducted from March 12th 2012 to June 4th 2016, and data analysis was conducted from June 5th 2016 to April 18th 2017.

Data sharing

With the publication of this article, the full dataset will be freely available online in Mendeley Data, a secure online repository for research data, which allows to archiving any file type and assigns a permanent and unique digital object identifier (DOI) so that the files can be easily referenced (DOI to be added).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit for publication. ACi, TAF, GS, ACh, LZA and YO had full access to all the data, and ACi was responsible for the decision to submit for publication.

RESULTS

Overall, 28,552 citations were identified by the search and 680 potentially eligible articles were retrieved in full text (Figure 1). We included 421 trials from the database search, additional 86 unpublished studies from trial registries and pharmaceutical company websites, and further 15 from personal communication or hand-searching other review articles. Overall, 522 double blind parallel RCTs (116,477 patients) conducted between 1979 and 2016, and comparing 21 antidepressants or placebo were included in the analysis (Figure 1, Appendix 1). Appendix 2 tabulates the characteristics of included studies. The mean study sample size was 248 participants (range 7 to 1162). Overall 87,052 participants were randomly assigned to an active drug and 29,425 to placebo. About two thirds of the sample population were female (62%) and the mean age was 44 years (range 29-80). The median duration of the acute treatment was 8 weeks (interquartile range [IQR] 6-8). Two-hundred and thirty six (47%) studies randomised participants to three or more arms and 304 (58%) were placebo-controlled. The majority of studies were multi-centre (391 of 472; 83%) and recruited outpatients only (335 of 437; 77%). Two-hundred and fifty two (57%) trials recruited patients from North America, 37 (8%) from Asia and 140 (31%) from Europe (79 trials [15%] did not specify). The great majority of patients had moderate to severe major depressive disorder, with a mean reported baseline severity score on Hamilton Depression Rating Scale 17-item of 25.7 (SD 3.97) among 464 out of 522 studies (89%). Response rate was imputed in about one fourth of cases (129 of 4734; 27%). Rescue medications (typically benzodiazepines or other sedative-hypnotics) were allowed in 187 studies (36%). Overall, 46 (9%) trials were rated as high risk of bias, 380 (72%) trials as moderate, and 96 (18%) as low (Appendix 3).

Four hundred and nine (78%) studies were funded by pharmaceutical companies. We retrieved unpublished information for 274 (52%) of the included trials. Consistent with the study protocol, the primary analysis was based on the 474 studies (106,966 patients) that used drugs within the licensed dose range (i.e. the dosage

approved by the regulatory agencies in USA and Europe - Appendix 4.4). The results of the primary analyses were replicated using *netmeta* in R.

Figure 2 shows the network of eligible comparisons for efficacy and acceptability (see Appendix 7.2 for the secondary outcomes). All antidepressant drugs, except milnacipran, had at least one placebo-controlled trial. Only levomilnacipran was not directly compared with at least another active drug in any of the networks. Appendix 7.1 provides detailed results of pairwise meta-analyses.

Figure 3 presents the network meta-analysis results for the primary outcomes. In terms of efficacy (432 RCTs, 102,443 patients), all antidepressants were more effective than placebo, with OR ranging between 2.13 (95% CrI 1.89 to 2.41) for amitriptyline to 1.37 (95% CrI 1.16 to 1.63) for reboxetine. In terms of acceptability (422 RCTs, 99,787 patients), agomelatine and fluoxetine were associated with fewer dropouts than placebo (OR 0.84, 95% CrI 0.72 to 0.97 and 0.88, 95% CrI 0.80 to 0.96, respectively); by contrast, clomipramine was worse than placebo (OR 1.31, 95% CrI 1.01 to 1.68). The relative efficacy of antidepressants compared to placebo is also demonstrated in remission (Appendix 7.2.4). The random-effects summary SMD for all antidepressants was 0.30 (95% CrI 0.26 to 0.34, p< 0.0001). In terms of dropouts due to adverse events, all active drugs were associated with higher withdrawal rates than placebo with OR ranging between 1.63 and 4.47, and 95% CrI excluding the null, except agomelatine (OR 1.21, 95% CrI 0.94 to 1.56) (Appendix 7.2.5).
In the analysis of response rate, 8% of the loops were inconsistent (17 out of 219 loops; p-value of the design by treatment test=0.063), and also 8% of the loops were inconsistent for dropouts (16 out of 210 loops; p-value=0.219). The median heterogeneity variances were estimated at 0.044 (0.028 to 0.063) and 0.040 (0.023 to 0.062) for response and dropout respectively, suggesting moderate to low heterogeneity.
Subgroup meta-regression analyses revealed that the use of placebo in trials was the strongest explanation of heterogeneity and inconsistency among those evaluated. Excluding placebo-controlled trials resulted in a 24% relative reduction in heterogeneity variance for response and 45% for dropout. Additionally, we found that smaller and older studies presented larger effects of the active interventions versus placebo (in particular for

amitriptyline, buproprion, fluoxetine and reboxetine) (See Appendix 7.6). The effect modification due to the use of placebo in study design was independent of year of randomisation or study precision (Appendix 8.5).

We also synthesised head to head studies separately to assess the differences between drugs. Figure 4 presents the table of primary outcomes (194 studies with at least two active arms at licensed dose and 34,196 patients). Agomelatine, amitriptyline, escitalopram, mirtazapine, paroxetine, sertraline, venlafaxine and vortioxetine were more effective than other antidepressants (ORs ranging between 1.12 and 1.96), while fluoxetine, reboxetine and trazodone were among the least efficacious drugs (ORs ranging between 0.51 and 0.89). In terms of acceptability, agomelatine, citalopram, escitalopram, fluoxetine, sertraline and vortioxetine were among the best drugs (ORs ranging between 0.42 and 0.81), while amitriptyline, clomipramine, duloxetine, fluvoxamine, reboxetine, trazodone and venlafaxine were the antidepressants associated with the highest dropout rates (ORs ranging between 1.23 and 2.37). Two-dimensional graphs about efficacy and acceptability in all studies and head-to-head studies only are reported in Figure 5. Results for the secondary outcomes were in line with the findings for the primary outcomes (see Appendix 8 for full details). Within the head-to-head comparisons, when a treatment was the novel/experimental one in the comparison it appeared to be significantly more effective than when that same treatment was the older/control one in the comparison (difference 1.18-fold (95% CI, 1.09- to 1.27-fold). Adjusting for this novelty effect diminished the differences between antidepressants.

We incorporated the GRADE judgments in Figure 4. The certainty of evidence for the relative treatment effects of efficacy and acceptability varied; it was moderate for most of the comparisons involving agomelatine, escitalopram, citalopram and mirtazapine, and low to very low for most comparisons involving vortioxetine, nefazadone, clomipramine, bupropion and amitriptyline (full details in Appendix 9). Appendix 10 presents the ranking of treatments based on cumulative probability plots and SUCRAs.

In accordance with the review protocol, we also carried out a sensitivity analysis including all the studies that used the drugs within the accepted doses (i.e. doses recommended in some international clinical guidelines – Appendix 4.4) and the results did not change materially (see Appendix 7.6.3.4).

DISCUSSION

The present study is based on 522 double-blind studies including 116,477 patients randomised to 21 individual first- and second-generation antidepressant drugs or placebo. The addition of the comparison with placebo is a major extension of our previous report.[7] The much larger evidence base (about 117,000 vs 26,000 patients), obtained through exhaustive search for published and unpublished information, allowed us to investigate additional important outcomes, such as remission, change in mood symptoms and dropouts due to side effects, and a number of methodological issues, such as sponsorship, dosing schedule, study precision and novelty effect.[15]

We found that all antidepressants included in the analysis were more efficacious than placebo in adult patients with major depressive disorder and the summary effect sizes were mostly modest. Some antidepressants, such as escitalopram, mirtazapine, paroxetine, agomelatine and sertraline had a relatively favourable balance in terms of response and dropout rate. By contrast, reboxetine, trazodone and fluvoxamine were associated with generally inferior efficacy and acceptability profiles, making them less favorable options. To make our results as relevant and robust as possible to inform clinical practice, we decided to focus on head-to-head studies and at the same time emphasise the certainty of the retrieved evidence. Our assessment overall found limited differences between antidepressants when all data were considered, while there was more diversity in the range of efficacy and drop out patterns seen across the head-to-head comparisons

The present findings in adults contrasts with the efficacy of antidepressants in children and adolescents, for whom fluoxetine is probably the only antidepressant that may reduce depressive symptoms.[21] This differential efficacy across age groups may reflect heterogeneous mechanisms and causes of depression,[22] the smaller number of studies in young people or different methodological issues affecting adult and paediatric trials.[23] The effect sizes were also smaller in more recent, and larger, placebo-controlled trials compared with older and smaller ones, which may be an indicator of bias.

Estimated differences between drugs were smaller in placebo-controlled trials compared to head-to-head studies. There are several potential explanations, as many factors have been associated with higher placebo response rates, such as randomization ratio and the expectation of receiving an active treatment, the therapeutic setting or the frequency of study visits.[24] In our dataset we found that response to the same antidepressant was on average smaller and dropouts more likely when a placebo arm was included in the trial. Moreover, for the same drug and the same probability of receiving placebo, larger all-cause dropout rates were associated with a lower response to treatment. The use of the last observation carried forward (LOCF) approach for imputing missing outcome data may have affected the estimates of treatment effect.[25] Depressive symptoms tend to spontaneously improve over time and this contributes to the high percentage of placebo responders in antidepressant trials.[26] Patients randomised to the active drug in a double-blind placebo-controlled trial may leave studies earlier than in head-to-head studies because they may suspect they have been allocated to placebo. Antidepressants usually take full effect only after weeks and so participants who dropout earlier tend to have poorer responses, which are carried forward to the end of the trial by the LOCF analysis. The final result can be an underestimate of the true efficacy of the active drug.

Another possible explanation could be a bias in conduct, analysis or reporting of head-to-head trials, driven by commercial interests.[27] In our analyses, funding by industry was not associated with significant differences in terms of response or dropout rates. However, non-industry funded trials were few and a number of trials did not report or disclose any funding. We also observed that drugs tended to show a more favourable efficacy

profile when they were novel and used as experimental treatments than when they had become old. This novelty effect may arise where a novel agent is perceived to be more effective and better tolerated; alternatively, selective analyses and outcome reporting bias may be more prominent when a treatment is first launched.[15]

Our literature search was as comprehensive as possible, including the largest amount of unpublished data to date, which are associated with less favourable effect sizes for antidepressants.[28] Although it is likely that a certain amount of unpublished data could not be retrieved, comparison-adjusted funnel plots were not suggesting that small studies give different results from larger studies either among placebo-controlled trials or among head-to-head comparison trials (Appendix 7.5). The estimates of treatment effect from our study are in line with previous reviews on the same matter,[28] but they are considerably more precise because of our larger quantity of data and resulting statistical power.

Our review has some limitations. According to GRADE framework, the quality of many comparisons was assessed as low or very low for amitriptyline, bupropion and venlafaxine, while it was often rated as moderate for agomelatine, escitalopram and mirtazapine. We incorporated the certainty of evidence in the main results of our analysis (Figure 4) to highlight the most robust findings for further use in clinical judgment. However, many trials did not report adequate information about randomisation and allocation concealment, which restricts the interpretation of these results. To increase the methodological rigour of the contributing evidence we included only double-blind trials, which were generally very similar in design and conduct. The poor information in terms of risk of bias assessment may be a matter of reporting, however we presented full details about the risk of bias of all included studies in Appendix 3. We did not conduct a formal cost-effectiveness analysis. All of the most effective antidepressants are now off patent and available in generic form. Some of the antidepressants are included in the WHO list of essential medicines, which makes them available worldwide and ready to use also in developing countries.

We analysed only average treatment effects and were not able to investigate potentially important clinical and demographical modifiers of treatment response at the individual patient level (for instance, age, gender, severity of symptoms, duration of illness). Patients recruited in randomised trials tend to be highly selected and we also excluded patients with psychotic or treatment resistant depression, which may limit the applicability of the results to these clinical subgroups, but it was intended as a methodological strength to assure transitivity in the network. We did not cover important clinical issues that may inform treatment decision making in routine clinical practice (for instance, specific adverse events, withdrawal symptoms or combination with non-pharmacological treatments) and also, due to the limited amount of information reported in the original studies, we were not able to quantify some outcomes such as global functioning. It should also be noted that some of the adverse effects of antidepressants occur over a longer period, meaning that positive results need to be taken with great caution, because the trials in this network meta-analysis were of short duration. The current report summarizes evidence of differences between antidepressants when prescribed as an initial treatment. Given the modest effect sizes, non-response to antidepressants will occur. Our information unfortunately cannot guide next-step choices after failure of such a first step, for which well-performed trials are scarce.[29]

Using the data made available in FDA and EMA website, the international trial registries and contacting study authors and pharmaceutical companies, we managed to incorporate in the analysis a considerable amount of unpublished data for some drugs (namely agomelatine, escitalopram, paroxetine, reboxetine, sertraline, venlafaxine, vilazodone and vortioxetine), but not for all the antidepressants included in the network meta-analysis. This limitation in the primary trials may affect the validity of the findings for some antidepressants, but the incorporation of both direct and indirect comparisons may have contributed to reduce the potential risk of bias.[30] We did our best to retrieve all unpublished data and contacted study authors for supplemental material, but we are aware that a significant proportion of the information is still not available to the public. There are online archives where trials are prospectively registered, however they collect reliable information only for the most recent studies and we cannot rule out the possibility that some studies are missing or the same study has been counted twice in our analyses. It is not uncommon for the same study to go by different names in different publications, which complicates the process of data synthesis.[31] Making the dataset fully and

freely available is our contribution to open science and we welcome any information that may help clarify any mistake in our dataset.

Notwithstanding these caveats, the findings from this comprehensive network meta-analysis represent the best currently available evidence-base to guide the choice about pharmacological treatment for acute major depressive disorder in adults. All statements comparing the merits of one antidepressant with another must be tempered by the potential limitations of the methodology,[32] the complexity of specific patient populations and the uncertainties that may result from choice of dose or treatment setting. We hope that these results will assist in shared decision-making between patients, carers and their clinicians.

References

1. Murray CJ, Barber RM, Foreman KJ, et al. Global, regional, and national disability- adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: quantifying the epidemiological transition. *Lancet* 2015; **386:** 2145-91.
2. GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; **388:** 1603-58.
3. World Health Organisation (WHO). Depression. Fact sheet (updated February 2017). http://www.who.int/mediacentre/factsheets/fs369/en/ (accessed April September 19th21st 2017).
4. Ioannidis JP. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008; **3:** 14.
5. Harmer CJ, Duman RS, Cowen PJ. How do antidepressants work? New perspectives for refining future treatment approaches. *Lancet Psychiatry* 2017; **4:** 409-18.
6. Higgins JPT, Welton NJ. Network meta-analysis: a norm for comparative effectiveness? *Lancet* 2015; **386:** 628-30.
7. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins J, Churchill R, Watanabe N, Nakagawa AN, Omori IM, McGuire H, Tansella M, Barbui C. Comparative efficacy and acceptability of 12 new generation antidepressants: a multiple treatment meta-analysis. *Lancet* 2009; **373:** 746-58.
8. Furukawa TA, Salanti G, Atkinson LZ, Leucht S, Ruhe HG, Turner EH, Chaimani A, Ogawa Y, Takeshima N, Hayasaka Y, Imai H, Shinohara K, Suganuma A, Watanabe N, Stockton S, Geddes JR, Cipriani A. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. *BMJ Open* 2016; **6:** e010919.
9. Hróbjartsson A, Thomsen AS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, Brorson S. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ* 2013; **185:** E201-11.

10. Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analyses. *Int Clin Psychopharmacol* 2005; **20:** 49-52.
11. Rhodes KM, Turner RM, Higgins JP. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *J Clin Epidemiol* 2015; **68:** 52-60.
12. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011; **64:** 163–71.
13. Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* 2012; **3:** 98-110.
14. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010; **29:** 932-44.
15. Salanti G, Dias S, Welton NJ, Ades AE, Golfinopoulos V, Kyrgiou M, Mauri D, Ioannidis JP. Evaluating novel agent effects in multiple-treatments meta-regression. *Stat Med* 2010; **29:** 2369-83.
16. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS One* 2013; **8:** e76654.
17. Lunn D, Spiegelhalter D, Thomas A, et al. The BUGS project: evolution, critique and future directions. *Stat Med* 2009; **28:** 3049–67.
18. Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: The network graphs package. *Stata Journal* 2015; **15:** 905-50.
19. Schwarzer G. Network meta-analysis. In: Schwarzer G, Carpenter JR, Rücker G, eds. Meta-analysis with R. Berlin, Heidelberg: Springer, 2015: 187–216.
20. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014; **9:** e99682.
21. Cipriani A, Zhou X, Del Giovane C, Hetrick SE, Qin B, Whittington C, Coghill D, Zhang Y, Hazell P, Leucht S, Cuijpers P, Pu J, Cohen D, Ravindran AV, Liu Y, Michael KD, Yang L, Liu L, Xie P. Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. *Lancet* 2016; **388:** 881-90.

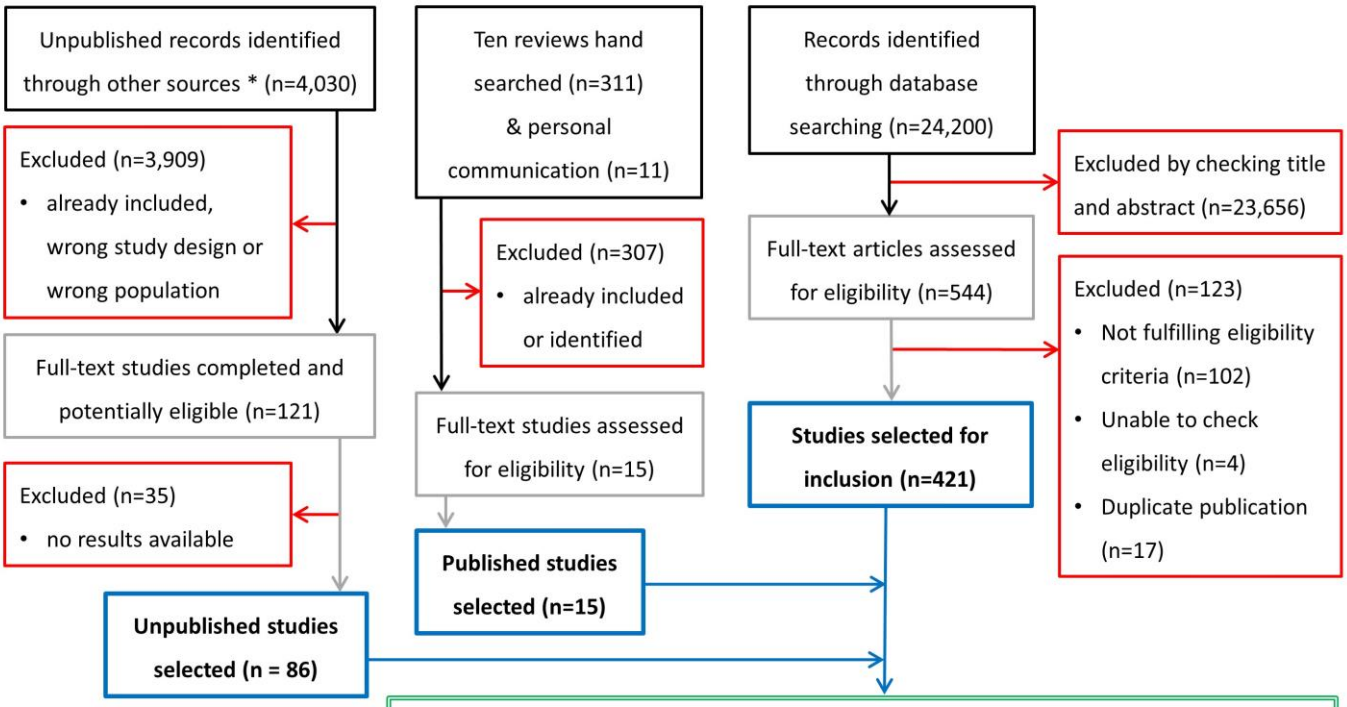
22. Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. *Lancet* 2012; **379:** 1056–67.
23. Walkup JT. Antidepressant Efficacy for Depression in Children and Adolescents: Industry- and NIMH-Funded Studies. *Am J Psychiatry* 2017; **174:** 430-37.
24. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 2013; **170:** 723-33.
25. Cook RJ, Zeng L, Yi GY. Marginal analysis of incomplete longitudinal binary data: a cautionary note on LOCF imputation. *Biometrics* 2004; **60:** 820-28.
26. Furukawa TA, Cipriani A, Atkinson LZ, Leucht S, Ogawa Y, Takeshima N, Hayasaka Y, Chaimani A, Salanti G. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. *Lancet Psychiatry* 2016; **3:** 1059-66.
27. Perlis RH, Perlis CS, Wu Y, Hwang C, Joseph M, Nierenberg AA. Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *Am J Psychiatry* 2005; **162:** 1957-60.
28. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; **358:** 252-60.
29. Furukawa TA, Akechi T, Shimodera S, Yamada M, Miki K, Watanabe N, Inagaki M, Yonemoto N. Strategic Use of New generation antidepressants for Depression: SUN(^_^)D study protocol. *Trials* 2011; 12: 116.
30. Furukawa TA, Miura T, Chaimani A, Leucht S, Cipriani A, Noma H, Mitsuyasu H, Kanba S, Salanti G. Using the contribution matrix to evaluate complex study limitations in a network meta-analysis: a case study of bipolar maintenance pharmacotherapy review. *BMC Res Notes* 2016; **9:** 218.
31. Wager E. The need for trial identifiers. *Curr Med Res Opin* 2004; **20:** 203-06.
32. Ioannidis JP. Meta-analyses can be credible and useful: a new standard. *JAMA Psychiatry* 2017; **74:** 311-12.

Authors' contributions: ACi, TAF, GS and JRG conceived and designed the study. ACi, TAF, LZA, SL, HGR, YO, NT, YH, EHT, HI, KS and AT selected the articles and extracted the data. GS, Ach, JPTH and ME analysed the data. ACi, TAF, GS and JRG wrote the first draft of the manuscript. ACh, LZA, YO, SL, HGR, EHT, JPTH, ME and JPAI interpreted the data and contributed to the writing of the final version. All authors read and met the ICMJE criteria for authorship and agree with the results and conclusions of this Article.

Funding statement: NIHR Oxford Health Biomedical Research Centre.

Competing interests statement: TAF has received lecture fees from Eli Lilly, Janssen, Meiji, Mitsubishi-Tanabe, MSD and Pfizer and consultancy fees from Takeda Science Foundation. He has received research support from Mochida and Mitsubishi-Tanabe. SL has received honoraria for consulting/advisory boards from Alkermes, Eli Lilly, Janssen, Johnson & Johnson, Lundbeck, MedAvante, Roche, Otsuka and Teva; lecture honoraria from AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Janssen, Johnson & Johnson, Lundbeck (Institute), Pfizer, Sanofi-Aventis, ICON, AbbVie, AOP Orphan and Servier; for the preparation of educational material and publications from Lundbeck Institute and Roche; and Eli Lilly has provided medication for a trial with SL as the primary investigator. NT has received lecture fees from Otsuka and Meiji. YH has received lecture fees from Yoshitomi. All other authors report no conflict of interest.

Acknowledgements: This project was funded by the NIHR Oxford Health Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. ACi is supported by the NIHR Oxford cognitive health Clinical Research Facility. GS is a Marie Skłodowska-Curie fellow (MSCA-IF-703254). JRG is an NIHR Senior Investigator.



Total number of DB RCTs included in the network meta-analysis (n=522, N=116,477)

- Agomelatine vs placebo or another active comparison (n = 23)
- Amitriptyline vs placebo or another active comparison (n = 96)
- Bupropion vs placebo or another active comparison (n = 33)
- Citalopram vs placebo or another active comparison (n = 38)
- Clomipramine vs placebo or another active comparison (n = 20)
- Desvenlafaxine vs placebo or another active comparison (n = 9)
- Duloxetine vs placebo or another active comparison (n = 30)
- Escitalopram vs placebo or another active comparison (n = 42)
- Fluoxetine vs placebo or another active comparison (n = 117)
- Fluvoxamine vs placebo or another active comparison (n = 32)
- Levomilnacipran vs placebo or another active comparison (n = 6)
- Milnacipran vs placebo or another active comparison (n = 10)
- Mirtazapine vs placebo or another active comparison (n = 34)
- Nefazodone vs placebo or another active comparison (n = 21)
- Paroxetine vs placebo or another active comparison (n = 114)
- Reboxetine vs placebo or another active comparison (n = 17)
- Sertraline vs placebo or another active comparison (n = 54)
- Trazodone vs placebo or another active comparison (n = 26)
- Venlafaxine vs placebo or another active comparison (n = 68)
- Vilazodone vs placebo or another active comparison (n = 9)
- Vortioxetine vs placebo or another active comparison (n = 15)

Figure 1: Selection of included and excluded studies (with reasons). Black boxes present screened references; red boxes present excluded studies (with reasons); blue boxes present selected studies, and green boxes present studies included in the network meta-analysis. DB: double blind; RCTs: randomized controlled trials. * Industry websites, contact with authors and trial registries. Clinicaltrials.gov was searched by 'drug name' AND 'major depressive disorder' as the major heading. The total number of unpublished records is the total number of results doing this for each drug and on each unpublished database source. The main reasons for exclusion included open label/single blind studies, studies including patients with comorbid disorders and combination therapy trials. Searches were only conducted on completed trials, which also removed many ongoing/terminated results, especially from clinicaltrials.gov.

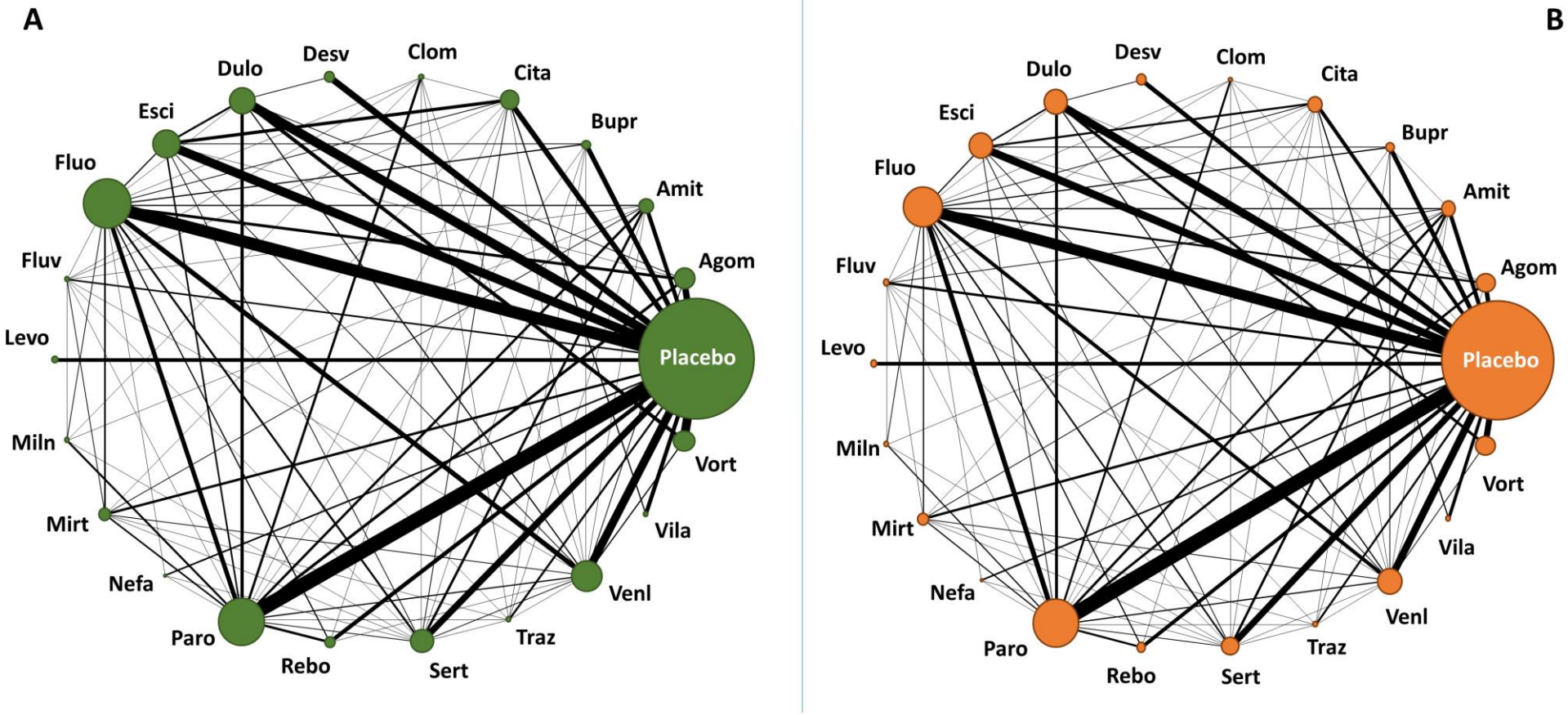


Figure 2: Network of eligible comparisons for efficacy (A) and acceptability (B). The width of the lines is proportional to the number of trials comparing every pair of treatments, and the size of every circle is proportional to the number of randomly assigned participants (sample size). Legend: Agom: agomelatine; Amit: amitriptyline; Bupr: bupropion; Cita: citalopram; Clom: clomipramine; Desv: desvenlafaxine; Dulo: duloxetine; Esci: escitalopram; Fluo: fluoxetine; Fluv: fluvoxamine; Levo: levomilnacipran; Miln: milnacipran; Mirt: mirtazapine; Nefa: nefazodone; Paro: paroxetine; Rebo: reboxetine; Sert: sertraline; Traz: trazodone; Venl: venlafaxine; Vila: vilazodone; Vort: vortioxetine.

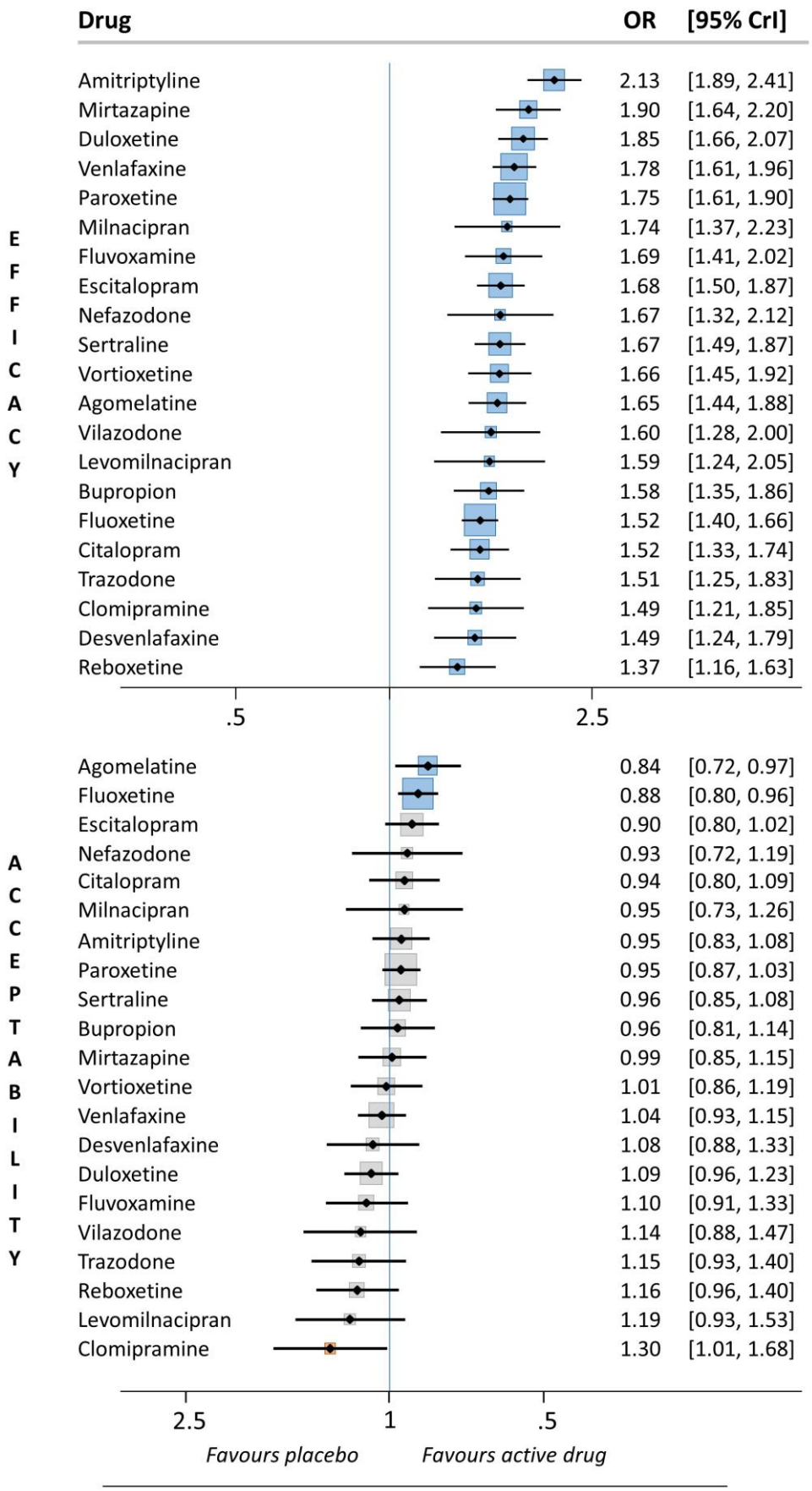


Figure 3: Forest plots of network meta-analysis results including all trials for efficacy (432 trials, 102,443 patients) and acceptability (422 RCTs, 99,787 patients) compared to placebo as reference compound. OR: Odds Ratio. CrI: Credible Interval.

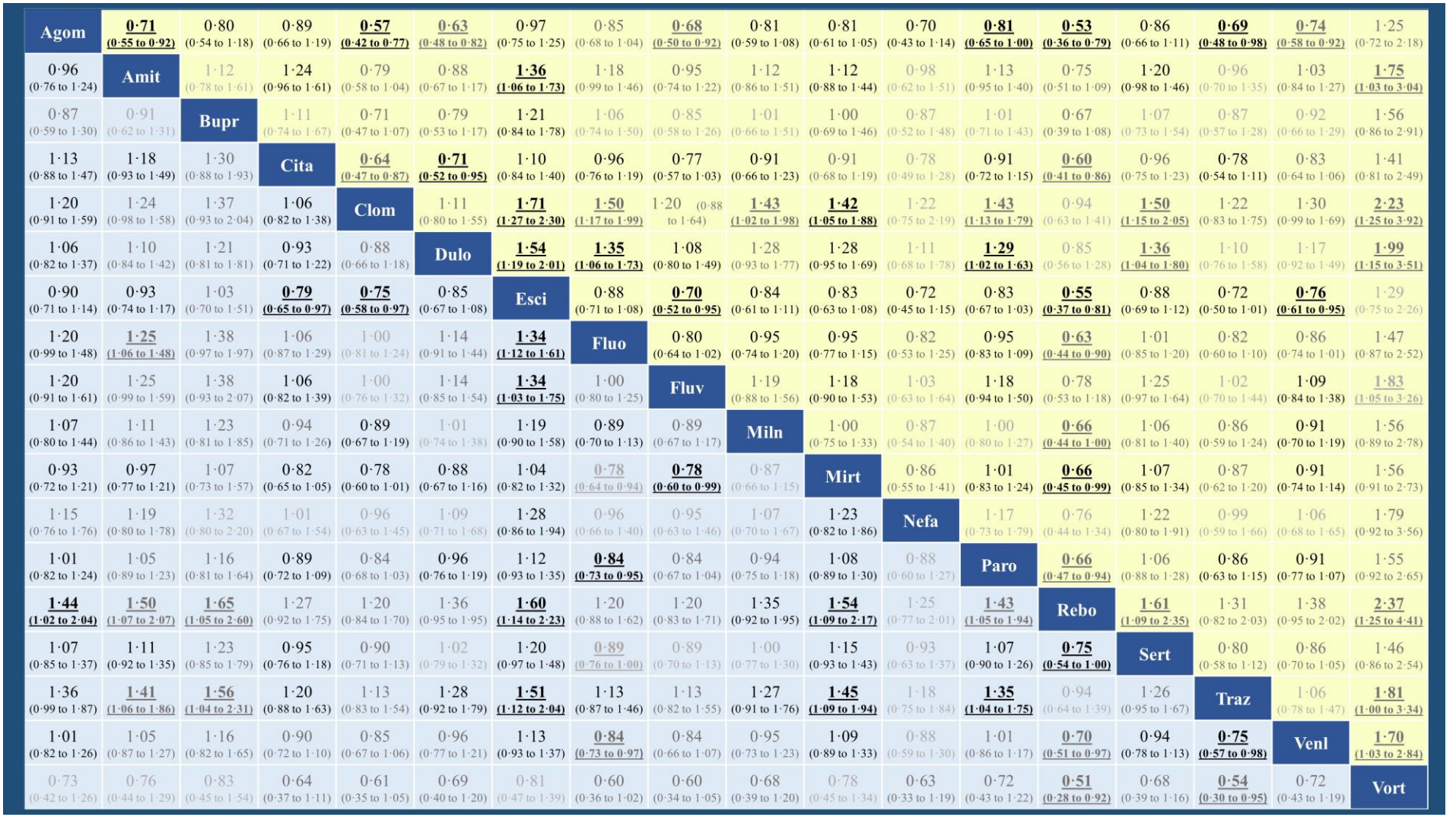


Figure 4: Efficacy (blue cells) and acceptability (yellow cells) of the 21 antidepressants. Drugs are reported in alphabetical order. Results are the ORs (with 95% CrI) in the column-defining treatment compared with the row-defining treatment. For efficacy, ORs higher than 1 favour the column-defining treatment (ie, the first in alphabetical order). For acceptability, ORs lower than 1 favour the first drug in alphabetical order. To obtain ORs for comparisons in the opposite direction, reciprocals should be taken. Significant results are in bold and underscored. The strength of the evidence (according to GRADE) was incorporated in this table (see main text and Appendix 9): figures reported in black represent "moderate quality"; figures reported in darkgray represent "low quality"; figures reported in light gray represent "very low quality" (see appendix 9 for full details about GRADE and reasons for downgrading). Agom=agomelatine. Amit=amitriptyline. Bupr=bupropion. Cita=citalopram. Clom=clomipramine. Dulo=duloxetine. Esci=escitalopram. Fluo=fluoxetine. Fluv=fluvoxamine. Miln=milnacipran. Mirt=mirtazapine. Nefa=nefazodone. Paro=paroxetine. Rebo=reboxetine. Sert=sertraline. Traz=trazodone. Venl=venlafaxine. Vort=vortioxetine. OR=Odds ratio. CrI=credibility interval.

Figure 5: Two-dimensional graphs about efficacy and acceptability in all studies (top) and head-to-head studies only (bottom). Individual drugs are represented by a coloured nodes, with corresponding confidence interval (bars). Legend. 1: agomelatine; 2: amitriptyline; 3: bupropion; 4: citalopram; 5: clomipramine; 6: desvenlafaxine; 7: duloxetine; 8: escitalopram; 9: fluoxetine; 10: fluvoxamine; 11: levomilnacipran; 12: milnacipran; 13: mirtazapine; 14: nefazodone; 15: paroxetine; 16: reboxetine; 17: sertraline; 18: trazodone; 19: venlafaxine; 20: vilazodone; 21: vortioxetine; 22: placebo.